**Information Retrieval**

1. **July 2014**

**Ex 1 [ranks 3+2]** Describe the dynamic indexing technique, assuming blocks of power-of-two number of documents, and compute the time complexity of each document insertion, assuming that the document processing takes constant time.

**Ex 2 [ranks 4+4]** Given a dictionary D of n stringsof total length N describe which data-structure solution you’d use to solve the following query on a string-pattern P:

* Find the strings of D which differ from P of one single character.
* Find the strings of D which differ from P of two characters.

For each solution, detail the query algorithm and its time/space complexity.

**Ex 3 [points 4+4]** Given symbols and probabilities: p(a) = 0.5, p(b)=p(c) = 0.25.

* Compress the string abca via Arithmetic coding
* Describe the iterative algorithm which represents a real number in binary, and comment on its correctness

**Ex 4 [points 3+3+3]** Define what is a strongly-connected component (SCC) in an directed graph G, how SCCs of G can be computed in RAM, and how they can be computed efficiently semi-externally.